Toward Automated Synthesis of Oligosaccharides and Glycoproteins

Pamela Sears and Chi-Huey Wong*

The discovery of previously unknown functions associated with carbohydrates and the study of their structure-function relations are of current interest in carbohydrate chemistry and biology. Progress in this area is, however, hampered by the lack of convenient and effective tools for the synthesis and analysis of oligosaccharides and glycoconjugates. Development of automated synthesis of such materials is necessary to facilitate research in this field. This review describes recent advances in carbohydrate synthesis, with particular focus on developments that have potential application to the automated synthesis of oligosaccharides, glycopeptides, and glycoproteins.

Saccharides have many key biological functions (*1*–*4*). When conjugated to protein to form glycoproteins, they can alter protein structure and function. As components of glycolipids, they can play pivotal roles in cell-cell recognition and signaling. The extracellular matrix contains proteoglycans, large glycoconjugates that not only modify the physicochemical properties of the solution but also are involved in many recognition processes. Although numerous carbohydrate structures occur in nature, in general, the role of saccharide structure in function has been minimally studied. This can be attributed mainly to the difficulty of synthesizing saccharides, especially when compared with proteins and nucleic acids. Nucleic acids can be made easily and cheaply via chemical and biological synthetic techniques, and protein sequences, which are encoded by DNA, can therefore be easily determined, produced, and manipulated through recombinant DNA technology. In addition, automatic synthesizers are available for the synthesis of polypeptides and oligonucleotides. Saccharides, however, are made (even in nature) in a seemingly haphazard way, with a diverse set of enzymes competing to produce very diverse products (*1*). There is no information carrier that "encodes" a particular saccharide structure, and so creating libraries of saccharides with methods akin to protein mutagenesis is not possible. Furthermore, unlike proteins and nucleic acids, saccharides are more difficult to synthesize because (i) the molecules are typically branched rather than linear, (ii) the monosaccharide units can be connected by α or β linkages, and (iii) oligosaccharide synthesis requires multiple selective protection and deprotection steps.

This last requirement is quite formidable, and currently there is no general route for saccharide synthesis. In a glycosidation reaction, both donors (monosaccharides activated for reaction) and acceptors (which receive the activated monosaccharide) contain many similar functional groups that must be differentiated and selectively protected. The product must then be selectively deprotected for the next round of reactions. The complexity of protecting-group manipulation increases with every additional glycosidic linkage. Development of stepwise solid-phase synthesis can simplify the intermediate work-up and purification steps, but the complexity of protecting-group manipulation remains the same. Because of this problem, there is currently no single stepwise synthetic approach that is applicable to the synthesis of all oligosaccharides or even just the >15 million possible tetrasaccharides that can be assembled from the nine common monosaccharides found in humans. In contrast, solid-phase synthesis of peptides and oligonucleotides involves only one protecting-group manipulation in the iterative process.

In the past few decades, however, the work of many research groups has started to open up new paths to saccharide and glycoconjugate synthesis. Coupling techniques with better yields and stereoselectivity have been worked out, and new protecting-group chemistries have also become available. The possibility of constructing libraries of saccharides, which was considered at one time to be a hopeless prospect, is now appearing to be feasible. The next step in making oligosaccharides widely accessible will be the automation of saccharide synthesis. This review focuses on the current state of the subject and emphasizes

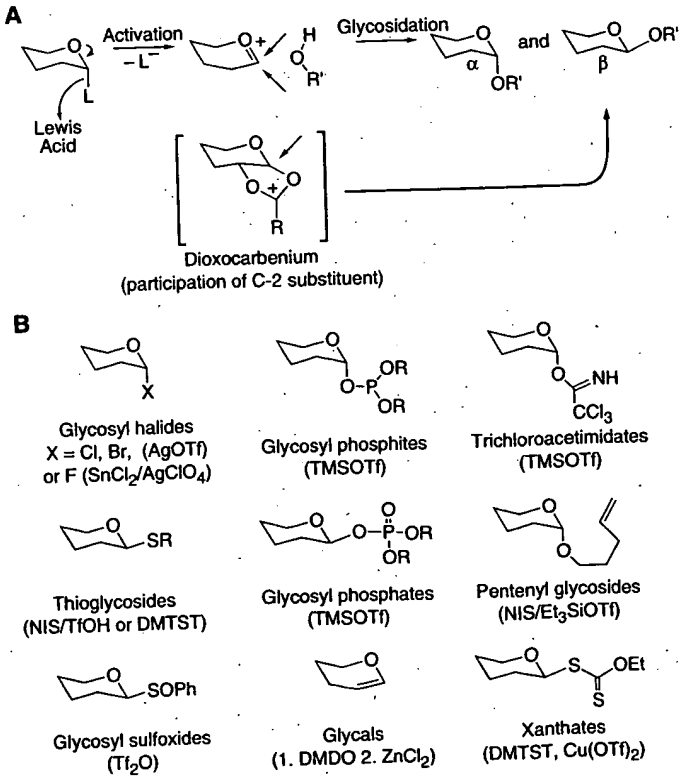

Fig. 1. (**A**) Common mechanisms for glycosidation. (**B**) Commonly used glycosidation reagents and their activators (in parentheses). Some of these glycosidation reagents can be used orthogonally. For example, the activator for glycosyl fluorides or phosphites will not activate thioglycosides or pentenyl glycosides. Abbreviations are as follows: DMDO, 3,3-dimethyldioxirane; DMTST, dimethylthiosulfonium triflate; Et, ethyl; L, leaving group; NIS, *N*-iodosuccinimide; R, variable group; OTf, triflate; TfOH, triflic acid; and TMSOTf, trimethylsilyl triflate.

Department of Chemistry and Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA.

*To whom correspondence should be addressed. E-mail: wong@scripps.edu

the developments with potential application to the automated synthesis of saccharides, glycopeptides, and glycoproteins.

Chemical Synthesis of Oligosaccharides

Several approaches have been taken with success for the chemical synthesis of oligosaccharides (Fig. 1) (*5–24*). Most involve the activation of the anomeric leaving group with a Lewis acid and then displacement of that leaving group by the free hydroxyl of the acceptor sugar. The Koenigs-Knorr method of coupling glycosyl halides, one of the first techniques to gain widespread usage, is still in common use (*5*), and most other glycosidation reagents used to date proceed by the same basic mechanism. The relative instability of the sugar halide necessitates the construction of the saccharide from the reducing end, and in fact, many of the most successful approaches are those that minimize side reactions of the activated sugar. New leaving groups have been further developed to improve the stability of the glycosyl donor and their reactivity. Trichloroacetimidates (*6*), prepared by the reaction of free sugars with trichloroacetonitrile and base, are used most frequently for coupling, as are glycosyl sulfoxides (*7*), phosphites (*8, 9*), and phosphates (*10*) and thio- (*11*) and pentenyl glycosides (*12*). Another scheme for glycoside synthesis is to build the saccharide from the nonreducing to the reducing end with glycals (*13*), which can be activated through epoxidation for either direct attack of the epoxide with the aglycon or intermediate formation of, for example, the thioacetal or phosphate.

The control of anomeric configuration of the product can be complicated, especially because the reaction can occur readily via either an S_N1- or an S_N2-type reaction (first- or second-order nucleophilic substitutions, respectively). The anomeric configuration of the activated sugar therefore does not ensure the anomeric configuration of the product. Furthermore, which products form can be heavily influenced by the protecting groups used. Acyl protecting groups at C-2 can strongly direct the trans configuration at C-1 by forming an intermediate dioxocarbenium ion (Fig. 1A). In general, α-1,2-*cis*-glycosides, such as α-D-glucosides and α-D-galactosides, can be formed either by taking advantage of the kinetic anomeric effect (*14*) in the displacement of glycosyl halides and thioglycosides or by direct displacement of β-trichloroacetimidates under conditions that favor inversion (no participating substituent at C-2 and a nonpolar solvent) (*15*). β-1,2-*trans*-Glycosides, such as β-D-glucosides and galactosides, can be obtained by using polar media to favor S_N1 displacement and formation of the dioxocarbenium. Glucosyl and galactosyl phosphates have, in all cases explored, produced the β-1,2-*trans*-glycoside, regardless of the anomeric configuration of the phosphate (*10*), and glycal chemistry also produces mainly the β-anomer. α-1,2-*trans*-Glycosides, such as α-D-mannosides, are simple to obtain because they are favored both by the kinetic anomeric effect and by the presence of participating groups at C-2, but β-1,2-*cis*-glycosides are still quite difficult to construct. Preparation of the β-D-glucoside followed by inversion at C-2 has been one common method, and recent attempts to direct the attack of the incoming sugar by tethering it in a position that allows only β attack have been successful (*16–19*).

In general, control of anomeric stereochemistry is still a problem, especially when the neighboring group participation is lacking. Also, there are certain chemistries that do not work well with some sugars. In nature, only α–sialic acid linkages are observed, but sulfoxide and trichloroacetimidate chemistries only give the β-anomer, a problem that can be solved by using other activating groups such as phosphites (*8, 9*), thioglycosides (*20*), and 2-xanthates (*21*).

In automating oligosaccharide synthesis, it is most convenient for the reactions to be performed on solid phase. This approach allows the rapid removal of reactants, relatively easy purification, and (in the case of library construction) the encoding of the product either by position (as in a two-dimensional array "chip" format) or, for "mix and split"–type library construction, by an accessory encoding reaction (*7*), in which labels are added to the solid support as the chain is extended or by radio frequency–encoded combinatorial chemistry technology (*22*). Most of the saccharide-synthesis techniques outlined above have been applied to solid-phase synthetic strategies on a variety of supports (*7, 13, 22–26*). Polystyrene-based resins, such as the Merrifield resin, are commonly used (*6, 24*), although these do not necessarily have the optimal characteristics for the synthesis of sugars with regard to swelling properties and reactant accessibility, particularly in hydrophilic media (*26*). More hydrophilic supports, such as polyethylene glycol–based resins, have been used with good success (*26*), as have "hybrid" resins, such as Tentagel, that have a polystyrene core coated in polyethylene. To a lesser extent, soluble supports, such as polyethylene glycols and derivatives, have been used in oligosaccharide synthesis.

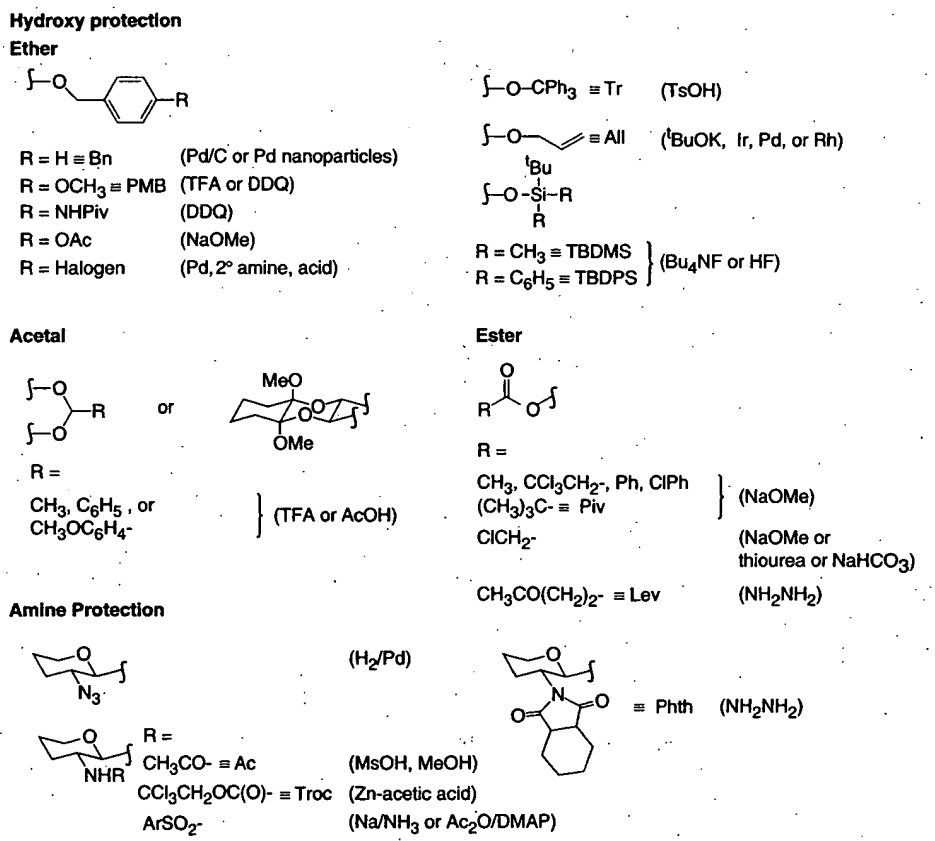

Fig. 2. Commonly used protecting groups and their removal conditions (in parentheses) [see (*5–23*) and citations therein]. Abbreviations are as follows: Ac, acetyl; All, allyl; Ar, aryl; Bn, benzyl; DDQ, 2,3-dichloro-5,6-dicyano-1,4-benzoquinone; DMAP, 4,-N,N-dimethylaminopyridine; Lev, levulinoyl; Me, methyl; MsOH, methanesulfonic acid; Ph, phenyl; Phth, phthalimidyl; Piv, pivaloyl; PMB, *p*-methoxybenzyl; TBDMS, *tert*-butyldimethylsilyl; TBDPS, *tert*-butyl-diphenylsilyl; ^{t}Bu, *tert*-butyl; TFA, trifluoroacetic acid; Tr, trityl; Troc, trichloroethoxycarbonyl; and TsOH, *p*-toluenesulfonic acid.

There are many disadvantages to using a solid support, however. Oligosaccharides and glycopeptides are sterically hindered compounds. Blocking one side of the molecule further with a solid support is likely to drop yields dramatically. Long flexible linkers can be used to alleviate this problem somewhat, but such linkers must be both cleavable and yet compatible with the coupling and protection-deprotection reactions [e.g., photo- or enzyme-sensitive linkers or linkers that can be cleaved by Pd(0) or by olefin cross metathesis]. Monitoring the reaction progress on solid phase is also not trivial. In addition, protecting-group manipulation on resins is extremely difficult. Heterogeneous catalysts, such as palladium supported on carbon, are not effective in solid-phase synthesis because of mass transport and surface-contact problems. Palladium nanoparticles, however, were found to be useful in the debenzylation of sugars attached to a polyethylene glycol–acrylamide resin (*27*).

The most challenging task, however, is the selection of orthogonal protecting groups and their selective manipulation during synthesis. Commonly used protecting groups include benzyl or silyl ethers and derivatives, as well as acid- or base-sensitive protecting groups (*15, 23, 28*) (Fig. 2). Although conditions have been developed for their selective deprotection, in general, their application to the synthesis of oligosaccharide libraries with great diversity has not been demonstrated. To date, the largest oligosaccharide made by solid-phase synthesis is that reported by Nicolaou (*25*) and Seeberger (*29*). Both groups synthesized the same branched dodecasaccharide on solid phase by using phenyl thioglycosides (*25*) or glycosyl phosphates and imidates (*29*), and the products were released from the support with photolysis (*25*) or olefin cross metathesis (*29*).

Enzymatic Synthesis of Oligosaccharides

In the past few decades, enzymatic approaches have been gaining popularity for the synthesis of saccharides and glycopeptides (*30, 31*). Enzymes feature exquisite stereo- and regioselectivity and catalyze the reaction under very mild conditions. Extensive protection-deprotection schemes are thus unnecessary, and the control of anomeric configuration is simple. Glycosyltransferases, the enzymes that are naturally used to synthesize saccharides, and glycosidases, enzymes normally used to hydrolyze glycosidic bonds, have both been used. Drawbacks to an enzymatic approach are the availability and cost of the catalysts and substrates, which can be large. The enzymes themselves are in many cases only just becoming available, particularly in the case of glycosyltransferases. The substrates, which for glycosyltransferases are the nucleotide-activated sugars, are relatively expensive but can be prepared from sugars or sugar phosphates through enzymatic or biological methods that have been worked out (*30, 32*). Glycosidases, which use cheaper substrates (such as sugar halides and *p*-nitrophenyl glycosides), can be used, but the yields have typically been lower. However, the Withers group recently found that mutagenesis of glycosidases to remove one of the two catalytic carboxylates in the active site produces an enzyme, coined a "glycosynthase," that can catalyze the synthesis of a saccharide from a fluorosugar donor but cannot catalyze hydrolysis of the resulting product (*33*) (Fig. 3). Whether this approach will be applicable to other exo-glycosidases remains to be investigated.

Another drawback of the enzymatic approach is that although enzymes are excellent at catalyzing the synthesis of natural products, their ability to accept novel saccharides with unusual or unnatural sugars as substrates may be poor; at best, it will be unknown. Models for the substrate preferences of glycosyltransferases are currently unavailable, and alteration of their specificity with protein engineering has experienced limited success. Prediction of reaction products with novel substrates will become easier as the enzymes begin to enjoy more widespread use and their substrate specificities become better characterized. Since the preparative-scale enzymatic synthesis of *N*-acetyllac-

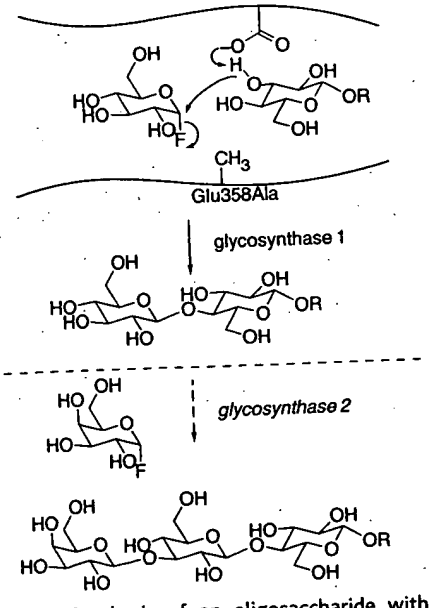

Fig. 3. Synthesis of an oligosaccharide with glycosynthases. In principle, exo-glycosidases can be genetically altered to accept glycosyl fluorides as donors to perform glycosidation.

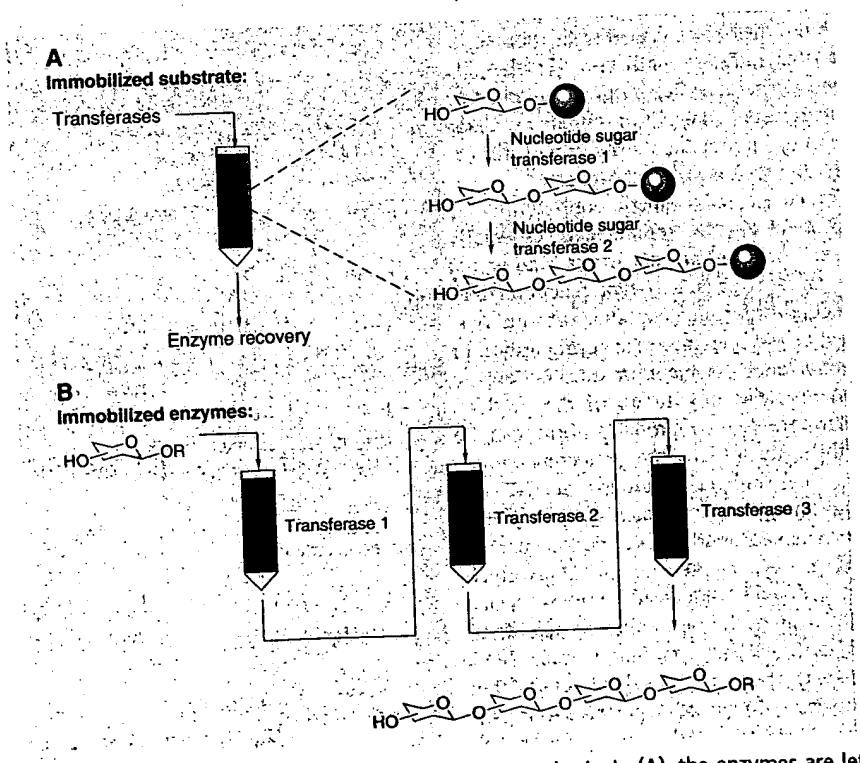

Fig. 4. Approaches to automated enzymatic saccharide synthesis. In (**A**), the enzymes are left in solution, and the growing saccharide is immobilized on the solid phase. This approach simplifies purification but requires an enzyme recovery step to avoid losing the expensive catalyst. Gray circles represent the solid support. In (**B**), the growing saccharide is attached to a water-soluble polymer, which is passed across columns of immobilized enzymes. Product recovery at the end can be accomplished by precipitating the polymer or by affinity techniques if the polymer is tagged with an affinity ligand (such as biotin). Alternatively, both enzymes and substrates can be used in free form.

tosamine involving sugar nucleotide regeneration in the 1980s (*34*), enzymatic approaches have been used in the synthesis of a great number of oligosaccharides and glycoconjugates (*31*). Further improvement in the area with the multiple enzymes required for sugar nucleotide regeneration immobilized on beads has been developed (*35*).

The application of enzymes to an automated scheme is possible. The logic of such a reaction scheme is conceptually simple because it is determined by the enzymes' preferred reaction: The saccharide must be built stepwise, in a linear fashion, from the reducing end (Fig. 4). Conducting the reaction on solid phase will require supplying the enzymes in solution, from which they must either be recovered for recycling or discarded. Recovery can be achieved via a variety of techniques, such as affinity-based capture (of affinity-tagged enzymes), passage through a microfilter, or enzyme precipitation. Enzymes are large molecules, and thus care must be used in choosing the support for solid-phase synthesis. The support, if porous, should have pores large enough to accommodate these macromolecules and should be hydrophilic to allow good swelling in water, or the support should be rigid so that the enzyme will not become entrapped (*30*). The use of long cleavable tethers to attach the growing saccharide may also help the substrate to enter the enzyme's active site. Many resins have been used, including polysaccharide-based resins (such as Sepharose), polyethylene-based resins (such as SPOCC), and polyacrylamide supports (*23, 26*). However, more standard solid-phase supports, such as derivatized silica and polystyrene, have also

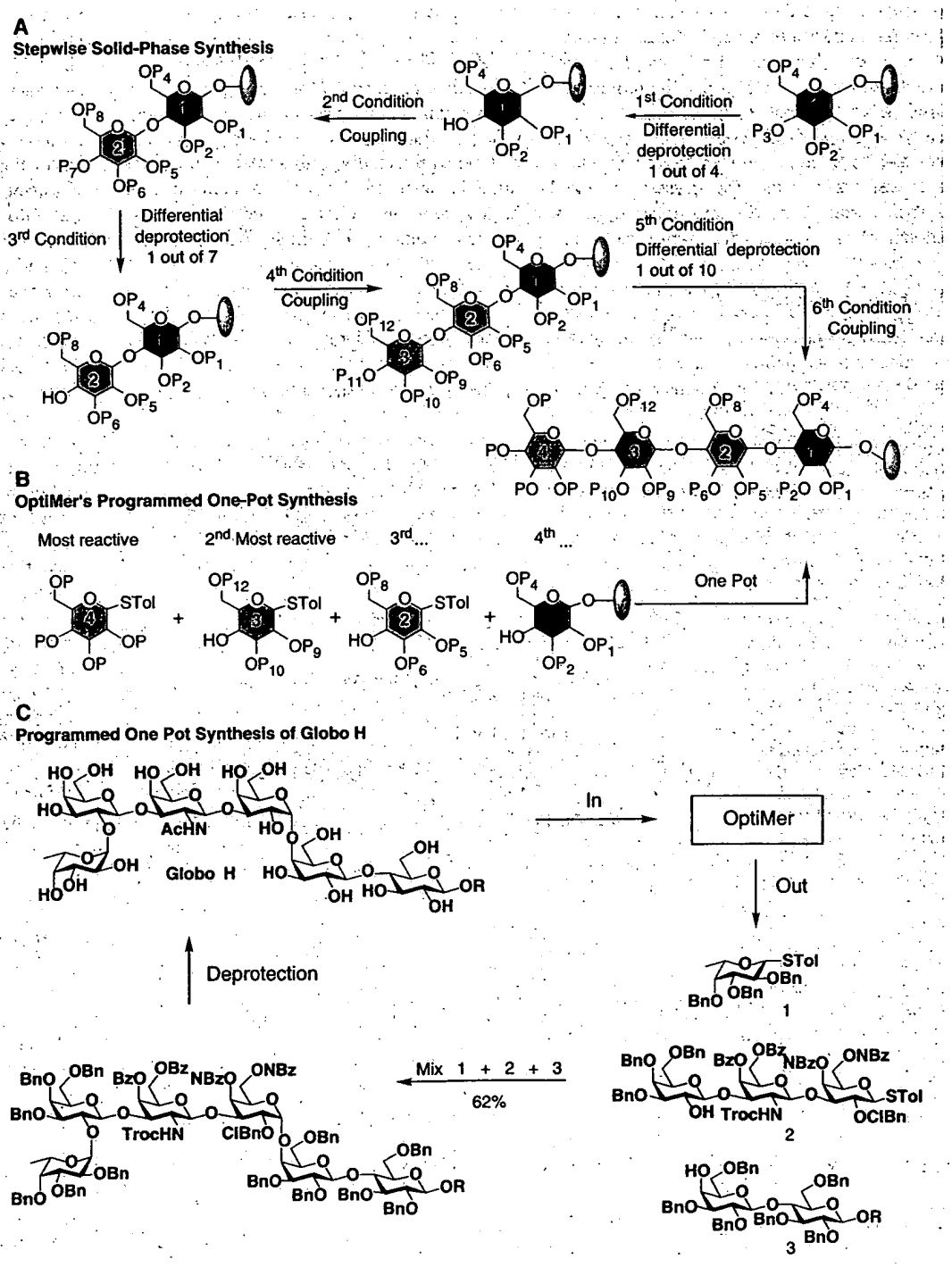

Fig. 5. (**A**) Traditional stepwise solid-phase synthesis requires on-resin protecting-group manipulation, which can become very complicated as the number of glycosidic linkages increases. (**B**) OptiMer's one-pot approach (*42*). OptiMer is a program that predicts the optimal type and order of addition of partially protected sugars, on the basis of a database of relative reactivities. This approach requires preparation of a number of building blocks with their glycosidation reactivities quantitatively measured. A reactivity difference of >1000 between the building blocks will give a high-yield coupling. No protecting-group manipulation and intermediate isolation is required during the one-pot synthesis. Red ovals in (A) and (B) represent the solid support. (**C**) Synthesis of the cancer antigen Globo H with OptiMer technology (*45*). In brief, the sequence of Globo H is entered into the computer, which predicts the best building blocks to be used. These building blocks are then mixed in sequence, starting with the most reactive one, in the presence of activator. The product obtained is then purified and deprotected to give the target. Abbreviations are as follows: Ac, acetyl; Bn, benzyl; Bz, benzoyl; Tol, tolyl; and Troc, trichloroethoxycarbonyl.

been used with success (*36, 37*). Solution-phase synthesis, although solving the problem of enzymatic accessibility, adds the problem of product recovery, which may be substantial, given the frequent complexity of the reaction buffer required for enzymatic reactions. A good approach may be to couple the substrate to a water-soluble polymer, which can be easily removed from solution either by precipitation of the polymer or by affinity-based capture (if an affinity label is attached to the support). Water-soluble supports, such as polyacrylamide that is not cross-linked, have been used in the enzymatic synthesis of saccharides and glycoconjugates, such as a pseudo-ganglioside named "pseudo-GM3" (*38*). Other water-soluble polymers, such as polyethylene glycol (*39*) and thermoresponsive polyacrylamide (*40*), may find use in enzymatic oligosaccharide synthesis. One can thus envision a scheme in which glycosyltransferases (and, if necessary, the enzymes required for the regeneration of its substrate) are immobilized onto a resin and packed into different columns. The substrate, free or bound to a water-soluble resin with an affinity tag, is passed through the columns in sequence, depending on the glycosyltransferases desired. Intermediate isolation, if necessary, can be achieved by capture of the substrate via affinity capture, for example. Attachment of enzymes and substrates to a support is, however, not trivial. In the end, given the high yields observed with glycosyltransferase-catalyzed glycosidations and the simplicity of product isolation, the choice of the reactor configuration is probably not critical, and both solution and solid-phase methods can be used for automated synthesis.

Programmable One-Pot Synthesis

A recent approach that shows promise for automation is the use of one-pot reaction schemes that use the reactivity profile of different protected sugars (*41, 42*) to determine the outcome. The reactivity of a sugar is highly dependent on the protecting groups and the anomeric activating group used. By adding substrates in sequence from most reactive to least reactive, one can assure the predominance of a desired target compound (Fig. 5). The key to this approach is to have extensive quantitative data regarding the relative reactivities of different protected sugars. A large amount of reactivity data for >100 protected *p*-methylphenyl thioglycosides was recently generated and used as the basis of a computer program (OptiMer) that selects the best reactants for one-pot synthesis of a target compound (*31, 42*). *p*-Methylphenyl thioglycosides were chosen because they are applicable to most monosaccharides and more reactive toward thiophilic activators, such as *N*-iodosuccinimide and dimethylthiosulfonium triflate, than are other thioglycosides (*11, 20, 43, 44*) that have been used in practical synthesis.

This approach has been used with success in the synthesis of a large number of oligosaccharides, including the cancer antigen Globo H hexasaccharide (Fig. 5C) (*45*). Further work is needed to design a complete set of building blocks (probably ~500 are needed) for use in the synthesis of most bioactive saccharides. So far, branch points have been incorporated by using the thioglycosides of disaccharides as reactants in the linear scheme. These reactions are typically performed in solution, but in order to facilitate removal of reactants at the end, the final acceptor may be attached to solid phase.

Future development in this approach is to expand the building block repertoire and to ensure their applicability in programmable one-pot synthesis. Compared with stepwise solid-phase synthesis, the one-pot approach requires protecting-group manipulation only at the stage of building block synthesis and thus holds greater potential for automation and for a greater diversity of oligosaccharide structures (Fig. 5).

Glycopeptide Synthesis

Attachment of saccharide chains to peptides can be accomplished in a number of ways. If the saccharide is built stepwise from the nonreducing to the reducing end, as is the case with glycal-based synthetic schemes and the one-pot strategy outlined above, then the ultimate acceptor can be an amino acid, peptide, or glycopeptide. For coupling to hydroxylated amino acids, such as serine or threonine, the chemistry is very much the same as that used to construct the glycosidic bonds: The activated anomeric position is directly attacked by a deprotected hydroxyl group on the peptide. In the case of NH_2-linked glycosides, the reducing-end sugar is typically prepared first as a sugar azide, which is then reduced and coupled to a free aspartate via carbodiimide activation. The acceptor can be an amino acid, for which the

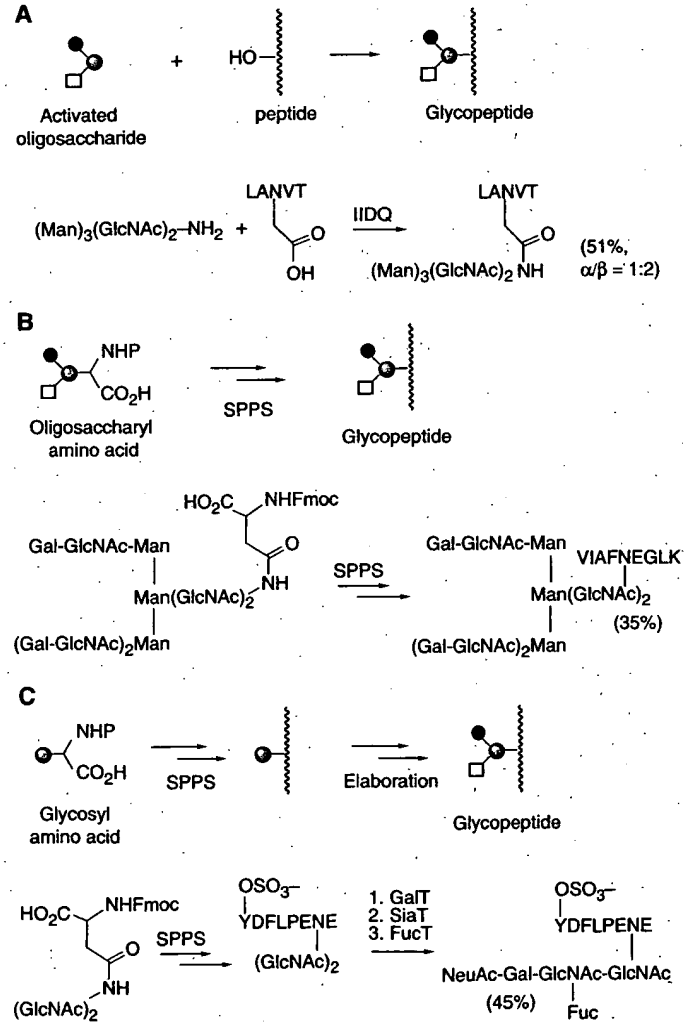

Fig. 6. Three approaches, which have the potential for automation, to preparing glycopeptides with complex glycans. (**A**) Convergent method for glycopeptide synthesis, in which the saccharide and peptide are built separately, then assembled at the end. Danishefsky's convergent synthesis of pentamannosyl pentapeptides is used. (LANVT, Lys-Ala-Asn-Val-Thr) (*50*). (**B**) The use of fully glycosylated amino acids in glycopeptide synthesis, along with the preparation of a nonapeptide (VIAFNEGLK, Val-Ile-Ala-Phe-Asn-Glu-Gly-Leu-Lys) with triantennary NH_2-linked glycan via this approach (*51*). (**C**) Solid-phase synthesis of a simple glycopeptide (YDFLPENE, Tyr-Asp-Phe-Leu-Pro-Glu-Asn-Glu), followed by elaboration of the glycan. Synthesis of a sulfated fragment of the cell adhesion molecule PSGL-1 with a pendant sialyl Lewis x molecule is shown (*53*). Abbreviations are as follows: Fmoc, fluorenylmethoxylcarbonyl; Fuc, fucose; FucT, α-1,3-fucosyltransferase; Gal, galactose; GalT, β-1,4-galactosyltransferase; GlcNAc, *N*-acetyl glucosamine; IIDQ, 2-isobutoxy-1-isobutoxycarbonyl-1,2-dihydroquinoline; Man, mannose; NeuAc, *N*-acetyl neuraminic acid; and SiaT, α-2,3-sialyltransferase.

product can be incorporated into solid-phase peptide synthesis (SPPS) schemes to produce the target glycopeptide (*26, 46, 47*), or it may itself be the final polypeptide. Glycosylated amino acids bearing typically one to three sugars have been used successfully in solid-phase synthesis of many glycopeptides. Paulsen and co-workers (*48*) constructed a variety of octapeptides containing three glycosylation sites bearing either *N*-acetylgalactosamine (GalNAc) or β-galactosyl-1,4-GalNAc and have more recently synthesized mucin glycopeptides containing many common core structures (*49*).

One problem with attempting to couple a large saccharide to a polypeptide has been low yield, presumably due to steric factors. Notably, Danishefsky and co-workers coupled a high-mannose-type pentasaccharide to a variety of tripeptides in 50 to 60% yield (Fig. 6) (*50*). Likewise, attempting to use an amino acid with a large pendant oligosaccharide in a SPPS scheme will result in a rapid drop-off of coupling yields as the oligosaccharide increases in size and branching. This problem was apparent in the solid-phase syntheses of glycosylated nonapeptides of various sizes performed by Meldal and co-workers (*51*), who incorporated amino acids containing high-mannose-type oligosaccharides with as many as 11 sugars in a triantennary structure (Fig. 6). As the saccharide increased in size, yields of the product dropped from nearly 80% down to 35%. An alternate approach is to glycosylate the peptide in a stepwise fashion from the reducing to the nonreducing end through chemical or enzymatic methods. Typically, a singly glycosylated peptide is made by SPPS, the sugar is selectively deprotected, and the oligosaccharide is built up in a stepwise fashion. A chemoenzymatic synthesis of a glycopeptide derived from the mucosal addressin cell adhesion molecule 1 was performed in this manner, both in solution and on solid phase (*52*). The singly glycosylated peptide was constructed via SPPS, and the sugar was completely deprotected to provide the substrate for the action of three successive glycosyltransferases. Similarly, a fragment of the cell adhesion molecule PSGL-1 with an attached sialyl Lewis x oligosaccharide was built enzymatically onto a sulfated octapeptide with a pendant GalNAc (Fig. 6) (*53*). In principle, the strategies used in the automation of oligosaccharide synthesis should be applicable to the synthesis of glycopeptides, and the sugar chain can be further elongated, if necessary, through either chemical or enzymatic methods.

Glycoprotein Synthesis

Extension of glycosylated peptides into glycoproteins can be accomplished with a number of approaches. Short peptides can be coupled to larger ones by "native peptide ligation" strategies (*54*). An 82-residue glycoprotein with two sites of glycosylation was recently made through chemical peptide synthesis with glycosylated amino acids, followed by native peptide ligation (*55*). However, this strategy puts constraints on the amino acid sequence adjacent to the peptide bond to be formed; namely, the NH_2-terminal amino acid of the acyl acceptor must be cysteine. Other strategies are also being explored.

The simplest approach to glycoprotein preparation, of course, is to let a cell do the work and prepare glycoproteins via fermentation. Unfortunately, fermentation produces a population of many different glycoforms of a given protein (*1*) because the saccharide that a protein receives reflects the cumulative effort of many glycosidases and transferases, and the action of some of these will preclude the action of others. The glycan produced will be determined by many factors, including the local protein structure around the glycosylation site and the relative amounts of glycoprocessing enzymes produced in the cell. Many of these factors also vary with the cell line, so a glycoprotein produced in one cell line will have different glycosylation than the same protein produced in another cell line.

This mixture, however, can be used as a starting point in many schemes in which the sugar chain is digested down to a simple homogeneous core and then reelaborated enzymatically (Fig. 7). For example, N-glycosylated proteins can have the glycans digested down to the innermost *N*-acetylglucosamine by using endoglycosidases, thus converting a heterogeneous population to a homogeneous one in which each glycosylation site has only a single sugar attached. These simple glycoproteins can then be elaborated enzymatically to increase the size and complexity of the glycan by using glycosyltranferases (*56*) or endoglycosidase-catalyzed transglycosylation (*57, 58*). The transglycosidase approach is limited by the substrate specificity of the endoglycosidases, which are enzymes that cleave between the innermost *N*-acetylglucosamine residues of

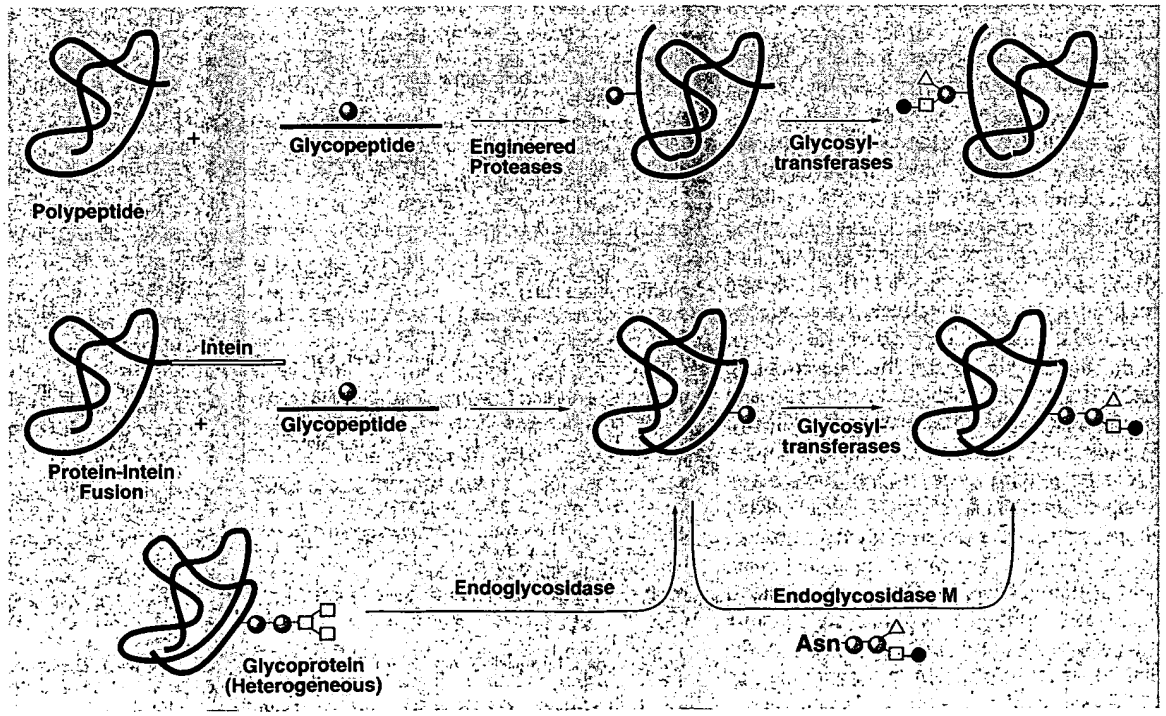

Fig. 7. Biochemical approaches to homogeneous glycoprotein preparation. The glycopeptide used can be prepared by solid phase and coupled to the NH_2-terminus of a truncated protein with protease or used to replace the COOH-terminal intein of a protein-intein fusion. Alternatively, the sugar chain of glycoproteins prepared by fermentation can be remodeled with enzymes. All proteins used can be prepared using recombinant DNA technology and fermentation.

N_2-linked oligosaccharides. Although many of the endoglycosidases are quite specific for certain classes of N_2-linked glycans, one that shows excellent promise is endoglycosidase M from *Mucor hiemalis*, which accepts a wide range of high-mannose–, hybrid-, and complex-type glycans (*58*). The disadvantage of these approaches is that the glycosylation site is limited to the one produced by the particular cell line used for production. Another option is to remove the glycosylated sections by using proteases and then reattach short, chemically synthesized glycopeptides in their place (*56*). This ligation can be accomplished enzymatically through the use of proteases or inteins, self-splicing polypeptides that are able to excise themselves from proteins posttranslationally. In the latter case, the peptide segment to be replaced is substituted at the genetic level with the sequence encoding the intein.

Proteases can catalyze peptide synthesis using either the thermodynamic approach or the kinetic approach. In the thermodynamic approach, peptides are condensed to form the larger product typically by precipitation of the product or by conducting the reaction in a solvent with low water activity. A more useful approach, as far as enzyme activity, stability, and solubility are concerned, is the kinetic approach, in which a peptide ester undergoes a competition between hydrolysis and aminolysis. The ratio of aminolysis to hydrolysis can be improved by adding an organic cosolvent to lower the water concentration and suppress amine ionization, by increasing the amine nucleophile concentration, or by modifying the enzyme active site. With regard to enzyme modification, the conversion of the active-site serine of serine proteases to a cysteine has been shown to be highly effective for creating a peptide ligase (*59, 60*). Glycosylation of proteins has long been known to render them less susceptible to protease activity, and so it might be inferred that glycopeptides would be difficult to couple using proteases. A systematic study of subtilisin-catalyzed synthesis of glycopeptides showed that the protease could couple glycopeptides successfully, provided that the glycosylation site was not at the forming bond and that the coupling yields improved as the glycosylation site was placed farther away from it (*37, 60*). The most effective and practical glycopeptide ester leaving group is the benzyl-type ester generated from a modified Rink amide resin and cleaved with trifluoroacetic acid (*37*).

An alternate approach is to use intein-mediated coupling of glycopeptides to larger proteins. It is possible to intervene in the natural splicing reaction by removing the COOH-terminal extein, then allowing the reaction to be completed with an exogenously added nucleophile (*61*), which may be a glycopeptide (*62*). As in the native peptide ligation strategy, the peptide must contain a cysteine at the NH_2-terminus.

Automatic synthesis of glycoproteins still represents a sizable challenge. However, the development of convenient methods for glycoprotein synthesis will allow us to study the effect of carbohydrates on glycoprotein structure and function, a subject of current interest (*63, 64*).

Future Prospects

Recent advances in synthetic carbohydrate chemistry have solved some major problems associated with carbohydrate research and have provided new strategies for tackling many interesting problems in glycobiology. Many technical problems that hinder the development of carbohydrate research still exist and remain to be solved. Of particular importance is the development of convenient and effective automated systems for the synthesis of oligosaccharides and glycoconjugates. Future efforts to reach this goal include, for example, the development of new methods for the rapid assembly of oligosaccharides and for the attachment of sugars to proteins, the design of new protecting groups, and the simplification of protecting-group manipulation. The development of such automated systems that are easily accessible to both biologists and chemists will have an important impact on our understanding of carbohydrates in biological systems and on the development of carbohydrate-based therapeutics.

References

1. P. Sears, C.-H. Wong, *Cell. Mol. Life Sci.* **54**, 223 (1998).
2. A. Varki, *Glycobiology* **3**, 97 (1993).
3. H. Lis, N. Sharon, *Eur. J. Biochem.* **218**, 1 (1993).
4. T. Kolter, K. Sandhoff, *Angew. Chem. Int. Ed. Engl.* **38**, 1532 (1999).
5. H. Paulsen, *Angew. Chem. Int. Ed. Engl.* **21**, 155 (1982).
6. J. Rademann, A. Geyer, R. R. Schmidt, *Angew. Chem. Int. Ed. Engl.* **37**, 1241 (1998).
7. R. Liang *et al.*, *Science* **274**, 1520 (1996).
8. T. J. Martin, R. R. Schmidt, *Tetrahedron Lett.* **33**, 6123 (1992).
9. H. Kondo, Y. Ichikawa, C.-H. Wong, *J. Am. Chem. Soc.* **114**, 8748 (1992).
10. S. Hashimoto, T. Honda, S. Ikegami, *J. Chem. Soc. Chem. Commun.* **1989**, 685 (1989).
11. R. J. Ferrier, R. W. Hay, N. Vethaviyasar, *Carbohydr. Res.* **27**, 55 (1973).
12. B. Fraser-Reid, J. R. Merritt, A. L. Handlon, C. W. Andrews, *Pure Appl. Chem.* **65**, 779 (1993).
13. S. J. Danishefsky, K. F. McClure, J. T. Randolph, R. B. Ruggeri, *Science* **260**, 1307 (1993).
14. R. U. Lemieux, K. B. Hendriks, R. V. Stick, K. James, *J. Am. Chem. Soc.* **97**, 4056 (1975).
15. S. L. Flitsch, G. M. Watt, in *Glycopeptides and Related Compounds*, D. G. Large, C. D. Warren, Eds. (Dekker, New York, 1997), pp. 207–243.
16. D. Crich, S. Sun, *J. Am. Chem. Soc.* **120**, 435 (1998).
17. F. Barresi, O. Hindsgaul, *J. Am. Chem. Soc.* **113**, 8376 (1991).
18. G. Stork, J. J. La Clair, *J. Am. Chem. Soc.* **118**, 247 (1996).
19. Y. Ito, T. Ogawa, *J. Am. Chem. Soc.* **119**, 5562 (1997).
20. P. J. Garegg, *Adv. Carbohyr. Chem. Biochem.* **52**, 179 (1997).
21. H. Lönn, K. Stenvall, *Tetrahedron Lett.* **33**, 115 (1992).
22. K. C. Nicolaou, X. Y. Xiao, Z. Parandoosh, A. Senyei, M. P. Nova, *Angew. Chem. Int. Ed. Engl.* **34**, 2289 (1995).
23. P. H. Seeberger, W.-C. Haase, *Chem. Rev.* **100**, 4349 (2000).
24. J. T. Randolph, K. F. McClure, S. J. Danishefsky, *J. Am. Chem. Soc.* **117**, 5712 (1995).
25. K. C. Nicolaou, N. Watanabe, J. Li, J. Pastor, N. Winssinger, *Angew. Chem. Int. Ed. Engl.* **37**, 1559 (1998).
26. P. M. St. Hilaire, M. Meldal, *Angew. Chem. Int. Ed. Engl.* **39**, 1162 (2000).
27. O. Kanie, G. Grotenberg, C.-H. Wong, *Angew. Chem. Int. Ed. Engl.* **39**, 4545 (2000).
28. B. Ernst, G. W. Hart, P. Sinay, Eds., *Carbohydrates in Chemistry and Biology* (Wiley-VCH, Weinheim, Germany, 2000), vol. 1.
29. O. J. Plante, E. R. Palmacci, P. H. Seeberger, *Science* **291**, 1523 (2001); published online 1 February 2001 (10.1126/science.1057324).
30. C.-H. Wong, R. L. Halcomb, Y. Ichikawa, T. Kajimoto, *Angew. Chem. Int. Ed. Engl.* **34**, 412 (1995).
31. K. M. Koeller, C.-H. Wong, *Chem. Rev.* **100**, 4465 (2000).
32. J. E. Heidlas, K. W. Williams, G. M. Whitesides, *Acc. Chem. Res.* **25**, 307 (1992).
33. L. F. Mackenzie, Q. Wang, R. A. J. Warren, S. G. Withers, *J. Am. Chem. Soc.* **120**, 5583 (1998).
34. C.-H. Wong, S. L. Haynie, G. M. Whitesides, *J. Org. Chem.* **47**, 5416 (1982).
35. X. Chen *et al.*, *J. Am. Chem. Soc.* **123**, 2081 (2001).
36. M. Schuster, P. Wang, J. C. Paulson, C.-H. Wong, *J. Am. Chem. Soc.* **116**, 1135 (1994).
37. K. Witte, O. Seitz, C.-H. Wong, *J. Am. Chem. Soc.* **120**, 1979 (1998).
38. K. Yamada, S.-I. Nishimura, *Tetrahedron Lett.* **36**, 9493 (1995).
39. S. P. Douglas, D. M. Whitfield, J. J. Krepinsky, *J. Am. Chem. Soc.* **113**, 5095 (1991).
40. D. E. Bergbreiter, J. Caraway, *J. Am. Chem. Soc.* **118**, 6092 (1996).
41. N. L. Douglas, S. V. Ley, U. Lücking, S. L. Warriner, *J. Chem. Soc. Perkin Trans. 1* **1998**, 51 (1998).
42. Z. Zhang *et al.*, *J. Am. Chem. Soc.* **121**, 734 (1999).
43. S. Hanessian, C. Bacquet, N. Lehong, *Carbohydr. Res.* **80**, C17 (1980).
44. K. C. Nicolaou, S. P. Seitz, D. P. Papahatjis, *J. Am. Chem. Soc.* **105**, 2430 (1983).
45. F. Burkhart, Z. Zhang, C.-H. Wong, *Angew. Chem. Int. Ed. Engl.*, in press.
46. H. Herzner, T. Reipen, M. Schultz, H. Kunz, *Chem. Rev.* **100**, 4495 (2000).
47. O. Seitz, *ChemBioChem* **1**, 214 (2000).
48. S. Peters, T. Biefeldt, M. Meldal, K. Bock, H. Paulsen, *Tetrahedron Lett.* **33**, 6445 (1992).
49. N. Mathieux, H. Paulsen, M. Meldal, K. Bock, *J. Chem. Soc. Perkin Trans. 1* **1997**, 2359 (1997).
50. S. J. Danishefsky, S. Hu, P. F. Cirillo, M. Eckhardt, P. H. Seeberger, *Chem. Eur. J.* **3**, 1617 (1997).
51. E. Meinjohanns, M. Meldal, H. Paulsen, R. A. Dwek, K. Bock, *J. Chem. Soc. Perkin Trans. 1* **1998**, 549 (1998).
52. O. Seitz, C.-H. Wong, *J. Am. Chem. Soc.* **119**, 8766 (1997).
53. K. M. Koeller, M. E. B. Smith, R.-F. Huang, C.-H. Wong, *J. Am. Chem. Soc.* **122**, 4241 (2000).
54. G. G. Kochendoerfer, S. B. H. Kent, *Curr. Opin. Chem. Biol.* **3**, 665 (1999).
55. Y. Shin *et al.*, *J. Am. Chem. Soc.* **121**, 11684 (1999).
56. K. Witte, P. Sears, R. Martin, C.-H. Wong, *J. Am. Chem. Soc.* **119**, 2114 (1997).
57. L.-X. Wang *et al.*, *J. Am. Chem. Soc.* **119**, 11137 (1997).
58. M. Mizuno *et al.*, *J. Am. Chem. Soc.* **121**, 284 (1999).
59. D. Y. Jackson *et al.*, *Science* **266**, 243 (1994).
60. P. Sears, M. Schuster, P. Wang, K. Witte, C.-H. Wong, *J. Am. Chem. Soc.* **116**, 6521 (1994).
61. C. J. Noren, J. Wang, F. B. Perler, *Angew. Chem. Int. Ed. Engl.* **39**, 450 (2000).
62. T. J. Tolbert, C.-H. Wong, *J. Am. Chem. Soc.* **122**, 5421 (2000).
63. R. A. Dwek, *Chem. Rev.* **96**, 682 (1996).
64. S. E. O'Connor, B. Imperiali, *Chem. Biol.* **55**, 427 (1998).

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☑ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☑ FADED TEXT OR DRAWING
- ☑ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.